Covering Numbers for Semicontinuous Functions

Johannes O. Royset
Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

Abstract. Considering the metric space of extended real-valued lower semicontinuous functions under the epi-distance, the paper gives an upper bound on the covering numbers of bounded subsets of such functions. No assumptions about continuity, smoothness, variation, and even finiteness of the functions are needed. The bound is shown to be nearly sharp through the construction of a set of functions with covering numbers deviating from the upper bound only by a logarithmic factor. The analogy between lower and upper semicontinuous functions implies that identical covering numbers hold for bounded sets of the latter class of functions as well, but now under the hypo-distance metric.

Keywords: covering numbers, metric entropy numbers, semicontinuous functions, epi-distance, Attouch-Wets topology, epi-convergence, epi-spline, approximation theory.

Date: April 29, 2016

1 Introduction

Covering numbers of classes of functions play central roles in parts of information theory, statistics, and applications such as machine learning; see for example [26, 16]. A large variety of results are available. The pioneering work [17, 11] deal with continuous and smooth functions; see [19] for a recent discussion. Functions of bounded variation are considered in [7] and analytic functions in [13]. An upper estimate for the covering numbers of the unit ball of Gaussian reproducing kernel Hilbert spaces is given in [29], with further refinements and applications in [27, 18]. Covering numbers of sets of convex functions are established in [14, 12], with significant improvements in [15].

In this paper, we provide upper bounds on the covering numbers of bounded subsets of extended real-valued lower semicontinuous (lsc) functions on $I\!R^d$ under the epi-distance metric. We permit any $d = 1, 2, ...$ and establish an upper bound on the ε-metric entropy number, which is the logarithm of the ε-covering number, that is of order $O(\varepsilon^{-d}(\log \varepsilon^{-1})^{d+1})$. This upper bound is nearly sharp as we construct a bounded set of lsc functions that has ε-metric entropy number $c\varepsilon^{-d} \log \varepsilon^{-1}$ for some $c > 0$.

It is well-known that bounded subsets of lsc functions are totally bounded under the epi-distance metric [20, Theorem 7.58] and consequently the covering numbers of such sets are finite. Here, we establish for the first time a quantification of these covering numbers. The class of lsc functions is quite

expansive as it includes functions defined on all of $I\!R^d$ that might even take on the values $\pm\infty$. This class is of interest in various function identification problems and their applications in statistics and operations research [25, 23, 24], and is fundamental to constrained minimization problems, which abstractly can be represented by lsc functions [20]. Since the negative of a lsc function is an extended real-valued upper semicontinuous (usc) function, the results developed here carry over directly to bounded subsets of the usc functions, now under the hypo-distance metric. Further applications arise in probability theory because there the hypo-distance metrizes weak convergence of distribution functions on $I\!R^d$, which obviously are usc [22]. Thus, as an example, the covering numbers given in this paper provide directly covering numbers for bounded sets of distribution functions on $I\!R^d$.

Technically, we rely on set-convergence of epi-graphs, coined epi-convergence by R. Wets in [28], which is quantified by a modified Pompeiu-Hausdorff distance between subsets of $I\!R^{d+1}$. This perspective was placed on a firm footing in [4, 2, 5, 8]; see also [6, 9, 10] for work on the convex case. The relevant results are available in [20, Chapter 7], which together with recent developments in [24, 21], provide the foundations for the present derivations.

After the review of background material in Section 2, the main theorems are stated in Section 3. Proofs and supporting results are given in Section 4.

2 Background

We let $\text{lsc-fcns}(I\!R^d) := \{f : I\!R^d \to \overline{I\!R} \;:\; f \text{ lsc and } f \not\equiv \infty\}$, where $\overline{I\!R} := I\!R \cup \{-\infty, \infty\}$. Thus, every $f \in \text{lsc-fcns}(I\!R^d)$ has a nonempty closed epi-graph $\text{epi}\, f := \{(x, x_0) \in I\!R^d \times I\!R \;:\; f(x) \leq x_0\}$. We adopt the sup-norm on $I\!R^d$, which leads to slight simplifications below, but other choices would only influence the constants in the main results. Let $I\!B(x, \rho) := \{y \in I\!R^d \;:\; \|x - y\|_\infty \leq \rho\}$, $\rho I\!B := I\!B(0, \rho)$, and $\rho\mathbb{S} := \rho I\!B \times [-\rho, \rho] \subset I\!R^d \times I\!R$. The epi-distance $d\!l$ is defined for any $f, g \in \text{lsc-fcns}(I\!R^d)$ as

$$d\!l(f, g) := \int_0^\infty d\!l_\rho(f, g) e^{-\rho} d\rho,$$

where the ρ-epi-distance, $\rho \geq 0$, is given by

$$d\!l_\rho(f, g) := \sup\left\{\left|\text{dist}\left(\bar{x}, \text{epi}\, f\right) - \text{dist}\left(\bar{x}, \text{epi}\, g\right)\right| \;:\; \bar{x} \in \rho\mathbb{S}\right\},$$

with dist giving the usual point-to-set distance, i.e., for $\bar{x} = (x, x_0) \in I\!R^d \times I\!R$,

$$\text{dist}\left(\bar{x}, \bar{C}\right) = \inf\left\{\max\{\|x - y\|_\infty, |x_0 - y_0|\} \;:\; (y, y_0) \in \bar{C}\right\} \text{ if } \bar{C} \subset I\!R^d \times I\!R \text{ is nonempty}$$

and $\text{dist}(\bar{x}, \emptyset) = \infty$. It is clear that $d\!l_\rho(f, g)$ is closely related to the Pompeiu-Hausdorff distance between $\text{epi}\, f$ and $\text{epi}\, g$, and in fact equivalent as ρ tends to infinity. Roughly speaking, the epi-distance between f and g is the weighted average of truncated versions of the Pompeiu-Hausdorff distance between their epi-graphs.

It is well-known that $(\text{lsc-fcns}(I\!R^d), d\!l)$ is a complete separable proper metric space [20, Theorem 7.58] (see also [24]). We recall that a metric space is proper if every closed ball in that space is compact.

The epi-distance induces the epi-topology on lsc-fcns($\mathbb{R}^d$), also called the Attouch-Wets topology. For $f^\nu, f \in$ lsc-fcns($\mathbb{R}^d$), $\nu \in \mathbb{N} := \{1, 2, ..., \}$,

$$d\!l(f^\nu, f) \to 0 \text{ if and only if } f^\nu \text{ epi-converges to } f.$$

Epi-convergence neither implies nor is implied by pointwise convergence. Uniform convergence ensures epi-convergence, but fails to handle extended real-valued functions satisfactory—a necessity in constrained optimization problems. Epi-convergence ensures convergence of solutions of minimization problems (see for example [20, Chapter 7] and [3, 1, 21]). It is therefore of particular importance in the area of optimization with numerous applications in machine learning, statistics, and control, but also of significance in study of partial differential equations where the closely related notion of Γ-convergence appears.

Following the usual definition of covering numbers, we let for any $F \subset$ lsc-fcns($\mathbb{R}^d$) and $\varepsilon > 0$, $N(F, \varepsilon; d\!l)$ be the smallest number of closed balls in lsc-fcns($\mathbb{R}^d$) with radii ε that cover F.

3 Main Results

In this section, we establish lower and upper bounds on the covering numbers for bounded subsets of lsc-fcns($\mathbb{R}^d$). The proofs are postponed to the subsequent section.

3.1 Theorem (covering numbers; upper bound) *Suppose that $d \in \mathbb{N}$ and $F \subset$ lsc-fcns($\mathbb{R}^d$) is bounded. Then, there exist $c \geq 0$ and $\bar{\varepsilon} > 0$ (independent of d) such that*

$$\log N(F, \varepsilon; d\!l) \leq \left(\frac{c}{\varepsilon}\right)^d \left(\log \frac{1}{\varepsilon}\right)^{d+1} \quad \textit{for all } \varepsilon \in (0, \bar{\varepsilon}].$$

The constant c depends on the size of a ball that contains the set under consideration, which brings in the need for boundedness. If the epi-distance had been defined using another norm on $\mathbb{R}^{d+1}$ than the sup-norm, c would have changed and possibly have depended on d. Bounded subsets of lsc-fcns($\mathbb{R}^d$) contain a wide variety of functions. For example, it follows from Proposition 4.1 below that the set $\{f \in \text{lsc-fcns}(\mathbb{R}^d) \;:\; f(0) \leq 0\}$ is contained in a ball centered at the zero-function with radius one. Thus, this set is bounded and can be covered as stipulated in Theorem 3.1. We observe that the significance of the point $0 \in \mathbb{R}^{d+1}$ derives from its selection as the center of the ball $\rho\mathbb{S}$ in the definition of $d\!l_\rho$. However, any other point could have been selected with only trivial implications.

Although a comparison to the classical result of $O(\varepsilon^{-d})$ for Lipschitz continuous functions on bounded subsets, which goes back to [17] (see for example [26, Theorem 2.7.1]), is not entirely relevant due the different metrics, we note that our bound is only slightly worse (a logarithmic term) for the larger class of lsc functions. Moreover, we do not require any bound on the variation of the functions and allow functions defined on all of $\mathbb{R}^d$, possibly extended real-valued.

The proof of Theorem 3.1 leverages recent approximation results for lsc functions. In [24] (see also [21]), we show that lsc functions can be approximated by piecewise constant functions called epi-splines that resemble the simple functions of integration theory. The error in approximation, in the epi-distance

metric, can be related directly to the number of pieces in the epi-splines. The challenge then becomes that of counting the number of balls centered at epi-splines that are needed to cover a particular subset of lsc-fcns($I\!R^d$).

We next state a lower bound on the covering numbers.

3.2 Theorem (covering numbers; lower bound) *For every $d \in I\!N$, there exist a bounded subset $F \subset$ lsc-fcns($I\!R^d$) and corresponding $c \geq 0$ and $\bar{\varepsilon} > 0$ (independent of d) such that*

$$\log N(F, \varepsilon; d\!l) \geq \left(\frac{c}{\varepsilon}\right)^d \log \frac{1}{\varepsilon} \text{ for all } \varepsilon \in (0, \bar{\varepsilon}].$$

In comparison with the upper bound of Theorem 3.1, we see that the lower bound differs by a logarithmic factor only. Let $\mathbf{0}$ be the function in lsc-fcns($I\!R^d$) that is identical to zero everywhere. We note that the size of the bounded set F in Theorem 3.2 does not have to be large. In fact, an examination of the proof reveals that F might be selected to have $d\!l(\mathbf{0}, f) \leq r$ for all $f \in F$, with $r > 1$ and arbitrarily close to 1.

The proof of Theorem 3.2 constructs a collection of functions which is finite on a grid of points in $[0, \rho]^d$, with $\rho > 0$ and grid points spaced roughly ε apart. At each of these grid points, a function takes on one value among a set of discretized values between 0 and ρ, again spaced roughly ε apart. Outside these grid points, the functions are infinity. It is clear that the number of such functions is $(\rho/\varepsilon)^n$, where $n = (\rho/\varepsilon)^d$. Thus, its logarithm is of the order $O(\varepsilon^{-d} \log \varepsilon^{-1})$. The proof proceeds by showing that no two of these functions are in a common ε-ball. Thus, it is necessary to have a number of balls to cover F that is at least the same as the number of functions constructed in this manner.

4 Proofs and Supporting Results

We start this section with estimates of the epi-distance. An auxiliary quantity is instrumental. For $\rho \geq 0$ and $f, g \in$ lsc-fcns($I\!R^d$), let

$$\hat{d\!l}_\rho(f, g) := \max\Big\{e\big(\operatorname{epi} f \cap \rho\mathbb{S}, \operatorname{epi} g\big),\ e\big(\operatorname{epi} g \cap \rho\mathbb{S}, \operatorname{epi} f\big)\Big\},$$

where the excess of a set C over a set D is given by

$$e(C, D) := \sup\{\operatorname{dist}(z, D) \ :\ z \in C\} \text{ if } C, D \text{ are nonempty},$$

$e(C, D) = \infty$ if C nonempty and D empty, and $e(C, D) = 0$ otherwise. Roughly speaking, $\hat{d\!l}_\rho(f, g)$ is the Pompeiu-Hausdorff distance between epi f and epi g, appropriately intersected with $\rho\mathbb{S}$. The relations among $d\!l$, $d\!l_\rho$, and $\hat{d\!l}_\rho$ are summarized next. The result is stated for the Euclidean norm on $I\!R^{d+1}$ in [20, Exercise 7.60], but remains unchanged in the present context of the sup-norm as established in [21].

4.1 Proposition *[21, 20, Exercise 7.60] For $f, g \in$ lsc-fcns($I\!R^d$) and $\rho \geq 0$, the following holds, where we use the notation $\delta_f = \operatorname{dist}(0, \operatorname{epi} f)$ and similarly for g:*

(i) $\hat{d\!l}_\rho(f, g) \leq d\!l_\rho(f, g) \leq \hat{d\!l}_{\rho'}(f, g)$ *for* $\rho' \geq 2\rho + \max\{\delta_f, \delta_g\}$;

(ii) $d\!l_\rho(f,g) \le \max\{\delta_f, \delta_g\} + \rho$;

(iii) $d\!l(f,g) \ge (1-e^{-\rho})|\delta_f - \delta_g| + e^{-\rho} d\!l_\rho(f,g)$;

(iv) $d\!l(f,g) \le (1-e^{-\rho}) d\!l_\rho(f,g) + e^{-\rho}[\max\{\delta_f, \delta_g\} + \rho + 1]$;

(v) $|\delta_f - \delta_g| \le d\!l(f,g) \le \max\{\delta_f, \delta_g\} + 1$.

The next result is essentially a direct consequence of Proposition 4.1.

4.2 Proposition *For* $f \in$ *lsc-fcns*$(\mathbb{R}^d)$ *and* $r \ge 0$, $d\!l(\mathbf{0}, f) \le r$ *implies* $\mathrm{dist}(0, \mathrm{epi}\, f) \le r$.

Proof. Proposition 4.1(v) gives that $r \ge d\!l(\mathbf{0}, f) \ge |\,\mathrm{dist}(0, \mathrm{epi}\,\mathbf{0}) - \mathrm{dist}(0, \mathrm{epi}\, f)| = \mathrm{dist}(0, \mathrm{epi}\, f)$. □

We rely on a "discretization" of lsc functions in terms of epi-splines [24, 21] and adopt the notation $\mathrm{cl}\, A$ for the closure of a subset A of a topological space. Moreover, for any $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ and $x \in \mathbb{R}^d$, let $\mathrm{liminf}_{x' \to x} f(x') := \lim_{\delta \downarrow 0} \inf_{x' \in \mathbb{B}(x,\delta)} f(x')$. Epi-splines are defined in terms a finite collection of subsets of $\mathbb{R}^d$. A finite collection $R_1, R_2, ..., R_K$ of open subsets of $\mathbb{R}^d$ is a partition of $\mathbb{R}^d$ if $\cup_{k=1}^K \mathrm{cl}\, R_k = \mathbb{R}^d$ and $R_k \cap R_l = \emptyset$ for all $k \ne l$. Specifically, an epi-spline $s : \mathbb{R}^d \to \mathbb{R}$, with partition $\mathcal{R} = \{R_k\}_{k=1}^K$ of $\mathbb{R}^d$, is a function that

$$\begin{aligned} &\text{on each } R_k,\ k = 1, ..., K, \text{ is constant real number,} \\ &\text{and for every } x \in \mathbb{R}^d, \text{ has } s(x) = \mathrm{liminf}_{x' \to x}\, s(x'). \end{aligned}$$

The family of all such epi-splines is denoted by e-spl$(\mathcal{R})$. The ability of epi-splines to approximate lsc functions is established by the next result; see [24, 21] for further information.

4.3 Proposition *For a partition* $\mathcal{R} = \{R_k\}_{k=1}^K$ *of* $\mathbb{R}^d$ *and* $\rho \ge 0$, *we have that for every* $f \in$ *lsc-fcns*$(\mathbb{R}^d)$, *there exists an* $s \in$ e-spl$(\mathcal{R})$ *such that*

$$\hat{d\!l}_\rho(s, f) \le \mu_\rho(\mathcal{R}) := \inf\big\{\eta \ge 0 \ :\ R_k \subset \mathbb{B}(x, \eta) \text{ for all } x \in \rho\mathbb{B} \text{ and } k \text{ satisfying } x \in \mathrm{cl}\, R_k\big\}.$$

If $\mu_\rho(\mathcal{R}) \le \rho$, *then* s *can be taken to satisfy* $-\rho' \le s(x) \le \max\{-\rho', \min\{\rho', f(x)\}\}$ *for any* $\rho' > \rho$ *and* $x \in \mathbb{R}^d$.

Proof. The first part of the proposition is a direct application of [21, Theorem 5.9]. The fact that s can be taken to satisfy $-\rho' \le s(x) \le \max\{-\rho', \min\{\rho', f(x)\}\}$ for any $\rho' > \rho$ follows from an examination of that theorem's proof. □

4.4 Proposition *[24, Theorem 3.17] If* $s, s' \in$ e-spl$(\{R_k\}_{k=1}^K)$, *then*

$$d\!l(s, s') \le \max_{k=1,...,K} \sup_{x \in R_k} |s(x) - s'(x)|.$$

We are then ready to give proofs of the main results.

Proof of Theorem 3.1. Since F is bounded, there exists an $r > 0$ such that $f \in F$ implies that $d\!l(\mathbf{0}, f) \leq r$. Let $\gamma_1, \gamma_2, \gamma_3 > 0$ be such that $\gamma_1 + \gamma_2 + \gamma_3 = 1$. Set $\bar{\varepsilon} \in (0, 1)$ such that

$$\frac{2(r+1)}{r}\left[\log\frac{1}{\varepsilon} + \log\frac{1}{\gamma_1} + \frac{r}{2} + \log(r+1)\right] - 1 > \gamma_2\varepsilon \text{ for all } \varepsilon \in (0, \bar{\varepsilon}].$$

Fix $\varepsilon \in (0, \bar{\varepsilon}]$ and define ρ to be the expression on the left-hand side of the previous inequality. We next construct a partition of $I\!R^d$ and set $\omega > 1$ and

$$n = \left\lceil \frac{2\omega\rho}{\gamma_2\varepsilon} \right\rceil,$$

where $\lceil a \rceil$ is the smallest integer no smaller than a. The partition is obtained by dividing the ball $I\!B(0, \omega\rho) = [-\omega\rho, \omega\rho]^d$ into n^d balls of equal size. Specifically, let $K = n^d + 1$ and $R_k, k = 1, 2, ..., n^d$, be the collection of nonoverlapping open boxes of the form $\prod_{i=1}^d (l_i^k, u_i^k)$, with $l_i^k, u_i^k \in I\!R$, $u_i^k - l_i^k = 2\omega\rho/n$, $l_i^k = 2(k-1)\omega\rho/n - \omega\rho$, $k = 1, ..., n$, and $\cup_{k=1}^{K-1} \operatorname{cl} R_k = [-\omega\rho, \omega\rho]^d$. Also, $R_K = I\!R^d \setminus [-\omega\rho, \omega\rho]^d$. We denote by $\mathcal{R} = \{R_k\}_{k=1}^K$ this partition. Clearly, $\mu_\rho(\mathcal{R}) = 2\omega\rho/n$. Next, we consider a discretization of parts of the range of lsc functions and set

$$m = \left\lceil \frac{\omega\rho}{\gamma_3\varepsilon} \right\rceil + 1.$$

The points $\sigma_j = -\omega\rho + 2(j-1)\omega\rho/(m-1)$, $j = 1, 2, ..., m$, discretize the interval $[-\omega\rho, \omega\rho]$. The epi-splines in e-spl($\mathcal{R}$) that take on one of these m values on each R_k is a collection of m^K unique epi-splines. Let $S \subset$ e-spl($\mathcal{R}$) be this collection of m^K epi-splines. That is, $s \in S$ if for every $k \in \{1, ..., K\}$, there exists a $j_k \in \{1, ..., m\}$ such that $s(x) = \sigma_{j_k}$ for $x \in R_k$. We now show that

$$\bigcup_{s \in S} I\!B(s, \varepsilon) \supset F.$$

Let $f \in F$ be arbitrary. By Proposition 4.3 and the fact that $\mu_\rho(\mathcal{R}) = 2\omega\rho/n \leq \gamma_2\varepsilon < \rho$, there exists $s_0 \in$ e-spl($\mathcal{R}$) such that

$$\hat{d\!l}_\rho(f, s_0) \leq \mu_\rho(\mathcal{R}) \text{ and } -\omega\rho \leq s_0(x) \leq \max\{-\omega\rho, \min\{\omega\rho, f(x)\}\} \text{ for } x \in I\!R^d.$$

Proposition 4.2 ensures that $\operatorname{dist}(0, \operatorname{epi} f) \leq r$. Thus, there exists an x such that $\|x\|_\infty \leq r$ and $f(x) \leq r$. Consequently, $s_0(x) \leq \max\{-\omega\rho, \min\{\omega\rho, f(x)\}\} \leq r$. So we also have that $\operatorname{dist}(0, \operatorname{epi} s_0) \leq r$.

Since $\varepsilon, \gamma_1 \leq 1$,

$$\rho \geq \frac{2(r+1)}{r}\left[\frac{r}{2} + \log(r+1)\right] - 1 = r + \frac{2(r+1)}{r}\log(r+1) \geq r.$$

Thus, using the notation $\bar{\rho} = (\rho - r)/2$, Proposition 4.1 gives that

$$\begin{aligned} d\!l(f, s_0) &\leq (1 - e^{-\bar{\rho}})d\!l_{\bar{\rho}}(f, s_0) + e^{-\bar{\rho}}(r + \bar{\rho} + 1) \\ &\leq \hat{d\!l}_\rho(f, s_0) + e^{-\bar{\rho}}(r + \bar{\rho} + 1) \\ &\leq \mu_\rho(\mathcal{R}) + e^{-\bar{\rho}}(r + \bar{\rho} + 1) \\ &= 2\omega\rho/n + e^{-\bar{\rho}}(r + \bar{\rho} + 1). \end{aligned}$$

In view of Proposition 4.4, there exists $s \in S$ such that $d\!\!l(s, s_0) \leq \omega\rho/(m-1)$ since we can select s such that $|s(x) - s_0(x)| \leq \omega\rho/(m-1)$ for all $x \in I\!\!R^d$. The triangle inequality then gives that

$$d\!\!l(f, s) \leq \omega\rho/(m-1) + 2\omega\rho/n + e^{-\bar{\rho}}(r + \bar{\rho} + 1).$$

It remains to show that the right-hand side is less than ε. We start with the last term. By concavity of the log-function, we have that

$$\log\left(\frac{1}{2}(\rho + r) + 1\right) \leq \log(r+1) + \frac{\rho - r}{2r+2}.$$

Consequently,

$$\begin{aligned}
\log\left[e^{-\bar{\rho}}(r + \bar{\rho} + 1)\right] &= \frac{1}{2}(r - \rho) + \log\left(\frac{1}{2}(\rho + r) + 1\right) \\
&\leq \frac{1}{2}(r - \rho) + \log(r+1) + \frac{\rho - r}{2r+2} \\
&= \frac{r}{2} - \frac{r(\rho+1)}{2(r+1)} + \log(r+1) \\
&= \log \gamma_1 \varepsilon,
\end{aligned}$$

where the last equality follows from inserting the expression for ρ. Thus, $e^{-\bar{\rho}}(r + \bar{\rho} + 1) \leq \gamma_1\varepsilon$. We then examine the second term. Inserting the expression for n, we obtain that

$$\frac{2\omega\rho}{n} \leq \gamma_2\varepsilon.$$

Finally, we consider the first term. In view of the definition of m, we have that

$$\frac{\omega\rho}{m-1} \leq \gamma_3\varepsilon.$$

Thus, $d\!\!l(f, s) \leq \varepsilon$. Since f is arbitrary, we have established that $\cup_{s \in S} I\!\!B(s, \varepsilon)$ covers F. The logarithm of the number of functions in S is $(n^d + 1)\log m$. At this point, the order of the result is immediate. A possible expression for the constant c is obtained as follows. Let $c_1 = 2(r+1)/r$ and

$$c_2 = \frac{2(r+1)}{r}\left[\log\frac{1}{\gamma_1} + \frac{r}{2} + \log(r+1)\right] - 1.$$

Thus, $\rho = c_1 \log \varepsilon^{-1} + c_2$. Moreover, let $c_3 = 2\omega/\gamma_2$ and $c_4 = \omega/\gamma_3$. Using these expressions, we find that

$$(n^d + 1)\log m \leq \left[\left(c_1c_3 + \frac{c_2c_3 + 1}{\log \bar{\varepsilon}^{-1}}\right)^d \left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)^d + 1\right] \log\left[\left(c_1c_4 + \frac{c_2c_4 + 2}{\log \bar{\varepsilon}^{-1}}\right)\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right]$$

Let

$$c_5 = c_1c_3 + \frac{c_2c_3 + 1}{\log \bar{\varepsilon}^{-1}} \text{ and } c_6 = c_1c_4 + \frac{c_2c_4 + 2}{\log \bar{\varepsilon}^{-1}}.$$

We then find that

$$(n^d+1)\log m \le c_7^d \left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)^d \left[\log c_6 + \log\frac{1}{\varepsilon} + \log\log\frac{1}{\varepsilon}\right], \text{ where } c_7 = c_5 + \frac{1}{\bar{\varepsilon}^{-1}\log\bar{\varepsilon}^{-1}}.$$

Using the fact that $\log\log\varepsilon^{-1}/\log\varepsilon^{-1} \le e^{-1}$ for $\varepsilon \in (0,1)$, we obtain that

$$(n^d+1)\log m \le c_7^d \left[\frac{\log c_6}{\log\bar{\varepsilon}^{-1}} + 1 + e^{-1}\right] \frac{1}{\varepsilon^d}\left(\log\frac{1}{\varepsilon}\right)^{d+1},$$

which gives a particular expression for c in the theorem statement. Since the choice of $\bar{\varepsilon}$ is independent of d, this c is independent of d. □

Proof of Theorem 3.2. Let $\rho > 0$ and $F = \{f \in \text{lsc-fcns}(\mathbb{R}^d) \ : \ f(x) \le \rho$ for at least one $x \in [0,\rho]^d\}$. We show that F cannot be covered with a lower number of balls than stipulated. Clearly, $\text{dist}(0, \text{epi}\, f) \le \rho$ for all $f \in F$. Thus, in view of Proposition 4.1(v), $d\!l(\mathbf{0}, f) \le \rho + 1$ for all $f \in F$ and F is therefore bounded.

Next, let $\varepsilon \in (0, \rho e^{-\rho}/6]$. We discretize $[0,\rho]^d$ by defining $x_i^k = k\rho/n_\varepsilon$, $k = 1, ..., n_\varepsilon - 1$ and $i = 1, ..., d$, where

$$n_\varepsilon = \left\lfloor \frac{\rho e^{-\rho}}{3\varepsilon} \right\rfloor \ge 2,$$

with $\lfloor a \rfloor$ being the largest integer not exceeding a. The discretization of $[0,\rho]^d$ then contains the points $(x_1^{k_1}, x_2^{k_2}, ..., x_d^{k_d})$, with $k_i \in \{1, 2, ..., n_\varepsilon - 1\}$ and $i = 1, ..., d$. Clearly, the distance between any two such points in the sup-norm is at least $\rho/n_\varepsilon \ge 3\varepsilon e^\rho$. We carry out a similar discretization of $[0,\rho]$ and define $y^l = l\rho/n_\varepsilon$, $l = 1, ..., n_\varepsilon$. The functions that are finite on the discretization points of $[0,\rho]^d$, with values at each such point equal to y^l for some l, and have value infinity elsewhere are given by S_ε, i.e.,

$$\begin{aligned} S_\varepsilon = \{&f \in \text{lsc-fcns}(\mathbb{R}^d) \ : \ \text{for each } x = (x_1^{k_1}, ..., x_d^{k_d}), \text{ with } k_i \in \{1, 2, ..., n_\varepsilon - 1\}, f(x) = y^l \\ &\text{for some } l = 1, ..., n_\varepsilon; f(x) = \infty \text{ otherwise}\}. \end{aligned}$$

Certainly, $S_\varepsilon \subset F$. We next define

$$G_\varepsilon(f) = \{g \in \text{lsc-fcns}(\mathbb{R}^d) \ : \ \hat{d\!l}_\rho(f,g) \le \varepsilon e^\rho\}, \ f \in \text{lsc-fcns}(\mathbb{R}^d).$$

We establish that $G_\varepsilon(f) \cap G_\varepsilon(f') = \emptyset$ for $f, f' \in S_\varepsilon, f \ne f'$. Suppose for the sake of a contradiction that there is a g with $g \in G_\varepsilon(f)$ and $g \in G_\varepsilon(f')$ for $f, f' \in S_\varepsilon$, $f \ne f'$. Then, $\hat{d\!l}_\rho(f,g) \le \varepsilon e^\rho$ and $\hat{d\!l}_\rho(f',g) \le \varepsilon e^\rho$. However, since $f \ne f'$, there exists a point $x \in [0,\rho]^d$ with $|f(x) - f'(x)| \ge 3\varepsilon e^\rho$. Without loss of generality, suppose that $f(x) \le f'(x) - 3\varepsilon e^\rho$. Since $f(z), f'(z) = \infty$ for all $z \ne x$ with $\|z - x\|_\infty < 3\varepsilon e^\rho$, we have that $\hat{d\!l}_\rho(f,g) \le \varepsilon e^\rho$ implies that $g(z) \le f(x) + \varepsilon e^\rho$ for some $z \in \mathbb{B}(x, \varepsilon e^\rho)$. Moreover, $\hat{d\!l}_\rho(f',g) \le \varepsilon e^\rho$ implies that $g(z) \ge f'(x) - \varepsilon e^\rho \ge f(x) + 3\varepsilon e^\rho - \varepsilon e^\rho = f(x) + 2\varepsilon e^\rho$ for all $z \in \mathbb{B}(x, \varepsilon e^\rho)$. Since this is not possible for g, we have reached a contradiction. Thus, $G_\varepsilon(f) \cap G_\varepsilon(f') = \emptyset$ for $f, f' \in S_\varepsilon, f \ne f'$.

By Proposition 4.1(i,iii), for any $f \in \text{lsc-fcns}(\mathbb{R}^d)$,

$$d\!l(f,g) \ge e^{-\rho} d\!l_\rho(f,g) \ge e^{-\rho}\hat{d\!l}_\rho(f,g) > e^{-\rho}\varepsilon e^\rho = \varepsilon \text{ for all } g \notin G_\varepsilon(f).$$

Hence, for $f \in S_\varepsilon$, an ε-ball that contains f needs to be centered at some $g \in G_\varepsilon(f)$. Since the sets $G_\varepsilon(f)$, $f \in S_\varepsilon$, are nonoverlapping, a cover of S_ε by ε-balls must involve a number of balls that is at least as great as the number of functions in S_ε, which is $n_\varepsilon^{m_\varepsilon}$, where $m_\varepsilon = (n_\varepsilon - 1)^d$. Thus,

$$\log N(F, \varepsilon; d\!l) \geq n_\varepsilon^d \log n_\varepsilon \geq \left(\frac{\rho e^{-\rho}}{3\varepsilon} - 2\right)^d \log\left(\frac{\rho e^{-\rho}}{3\varepsilon} - 1\right).$$

Let $c_1 = |\log(\rho e^{-\rho}/4)|$ and $\bar{\varepsilon} = \min\{\rho e^{-\rho}/12, e^{-2c_1}\}$. Continuing form the previous inequality, we then find that

$$\log N(F, \varepsilon; d\!l) \geq \left(\frac{\rho e^{-\rho}}{6}\right)^d \left[1 + \frac{\log(\rho e^{-\rho}/4)}{\log \varepsilon^{-1}}\right] \frac{1}{\varepsilon^d} \log \frac{1}{\varepsilon}.$$

Since $\log \varepsilon^{-1} \geq 2|\log(\rho e^{-\rho}/4)|$ for $\varepsilon \in (0, \bar{\varepsilon}]$, we have that

$$\log N(F, \varepsilon; d\!l) \geq \left(\frac{\rho e^{-\rho}}{6}\right)^d \frac{1}{2} \frac{1}{\varepsilon^d} \log \frac{1}{\varepsilon} \quad \text{for } \varepsilon \in (0, \bar{\varepsilon}],$$

and the conclusion is reached. $\square$

Acknowledgement. This work in supported in parts by DARPA under grant HR0011-14-1-0060.

References

[1] H. Attouch. *Variational Convergence for Functions and Operators*. Applicable Mathematics Sciences. Pitman, 1984.

[2] H. Attouch, R. Lucchetti, and R. J-B Wets. The topology of the ρ-Hausdorff distance. *Annali di Matematica pura ed applicata*, CLX:303–320, 1991.

[3] H. Attouch and R. J-B Wets. A convergence theory for saddle functions. *Transactions of the American Mathematical Society*, 280(1):1–41, 1983.

[4] H. Attouch and R. J-B Wets. Isometries for the Legendre-Fenchel transform. *Transactions of the American Mathematical Society*, 296:33–60, 1986.

[5] H. Attouch and R. J-B Wets. Quantitative stability of variational systems: I. The epigraphical distance. *Transactions of the American Mathematical Society*, 328(2):695–729, 1991.

[6] D. Aze and J.-P. Penot. Recent quantitative results about the convergence of convex sets and functions. In *Functional Analysis and Approximations. Proceedings of the International Conference Bagni di Lucca*, pages 90–110. Pitagora Editrice, 1990.

[7] P. L. Bartlett, S. R. Kulkarni, and S. E. Posner. Covering numbers for real-valued function classes. *IEEE Transactions on Information Theory*, 43(5):1721–1724, Sep 1997.

[8] G. Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer, 1992.

[9] G. Beer and R. Lucchetti. Convex optimization and the epi-distance topology. *Transactions of the American Mathematical Society*, 327(2):795–813, 1991.

[10] G. Beer and R. Lucchetti. The epi-distance topology: continuity and stability with application to convex optimization. *Mathematics of Operations Research*, 17:715–726, 1992.

[11] M. S. Birman and M. Z. Solomjak. Piecewise-polynomial approximation of functions of the classes w_p^α. *Mathematics of the USSR-Sbornik*, 73:295–317, 1967.

[12] E. M. Bronshtein. ε-entropy of convex sets and functions. *Siberian Mathematical Journal*, 17(3):393–398, 1976.

[13] A. Brudnyi. On covering numbers of sublevel sets of analytic functions. *Journal of Approximation Theory*, 162(1):72 – 93, 2010.

[14] R. M. Dudley. Metric entropy of some classes of sets with differentiable boundaries. *Journal of Approximation Theory*, 10(3):227–236, 1974.

[15] A. Guntuboyina and B. Sen. Covering numbers for convex functions. *IEEE Transactions on Information Theory*, 59(4):1957–1965, 2013.

[16] Y. Guo, P. L. Bartlett, J. Shawe-Taylor, and R. C. Williamson. Covering numbers for support vector machines. *IEEE Transactions on Information Theory*, 48(1):239–250, Jan 2002.

[17] A. N. Kolmogorov and V. M. Tikhomirov. Epsilon-entropy and epsilon-capacity of sets in functional spaces. *American Mathematical Society Translations, Series 2*, 17:277–364, 1961.

[18] T. Kühn. Covering numbers of Gaussian reproducing kernel Hilbert spaces. *Journal of Complexity*, 27(5):489–499, 2011.

[19] M. Pontil. A note on different covering numbers in learning theory. *Journal of Complexity*, 19(5):665–671, 2003.

[20] R.T. Rockafellar and R. J-B Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft*. Springer, 3rd printing-2009 edition, 1998.

[21] J. O. Royset. Approximations and solution estimates in optimization. In review. Preprint at http://faculty.nps.edu/joroyset/pubs.html; accessed April 29, 2016.

[22] J. O. Royset and R. J-B Wets. Variational theory for optimization under stochastic ambiguity. In review. Preprint at http://faculty.nps.edu/joroyset/pubs.html; accessed April 29, 2016.

[23] J. O. Royset and R. J-B Wets. From data to assessments and decisions: Epi-spline technology. In A. Newman, editor, *INFORMS Tutorials*. INFORMS, Catonsville, 2014.

[24] J. O. Royset and R. J-B Wets. Multivariate epi-splines and evolving function identification problems. *Set-Valued and Variational Analysis*, to appear, 2016. Available OnlineFirst.

[25] J.O. Royset and R. J-B Wets. Fusion of hard and soft information in nonparametric density estimation. *European J. of Operational Research*, 247(2):532–547, 2015.

[26] A. W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 2nd printing 2000 edition, 1996.

[27] J. Wang, H. Huang, Z. Luo, and B. Chen. Estimation of covering number in learning theory. In *Proceeding of the Fifth International Conference on Semantics, Knowledge and Grid, 2009*, pages 388–391, Oct 2009.

[28] R. J-B Wets. Convergence of convex functions, variational inequalities, and convex optimization problems. In R. Cottle, R. Giannessi, and J.-L. Lions, editors, *Varitional Inequalities and Complementary Problems*, pages 375–403. Wiley, New York, 1980.

[29] D.-X. Zhou. The covering number in learning theory. *Journal of Complexity*, 18(3):739–767, 2002.